Journal of Clinical and Translational Hepatology **2024** vol. 12(3) | 266–277
DOI: 10.14218/JCTH.2023.00457

Original Article

Prognostic Biomarkers for Hepatocellular Carcinoma Based on Serine and Glycine Metabolism-related Genes

Xufan Cai[1,2#], Fang Xu[3#], Zhaohong Wang[4], Hui Chen[4] and Shengzhang Lin[1*]

[1]Department of Clinical Medicine, School of Medicine, Hangzhou City University, Hangzhou, Zhejiang, China; [2]Zhejiang Chinese Medical University, Hangzhou, Zhejiang, China; [3]The First Affiliated Hospital of Bengbu Medical College, Bengbu, Anhui, China; [4]Department of Hepatobiliary and Pancreatic Surgery, The Second Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China

Received: 11 October 2023 | **Revised:** 26 December 2023 | **Accepted:** 2 January 2024 | **Published online:** 9 February 2024

Abstract

Background and Aims: Targeted therapy and immunotherapy have emerged as treatment options for hepatocellular carcinoma (HCC) in recent years. The significance of serine and glycine metabolism in various cancers is widely acknowledged. This study aims to investigate their correlation with the prognosis and tumor immune microenvironment (TIME) of HCC. ***Methods:*** Based on the public database, different subtypes were identified by cluster analysis, and the prognostic model was constructed through regression analysis. The gene expression omnibus (GEO) data set was used as the validation set to verify the performance of the model. The survival curve evaluated prognostic ability. CIBERSORT was used to evaluate the level of immune cell infiltration, and maftools analyzed the mutations. DsigDB screened small molecule compounds related to prognostic genes. ***Results:*** HCC was found to have two distinct subtypes. Subsequently, we constructed a risk score prognostic model through regression analysis based on serine and glycine metabolism-related genes (SGMGs). A nomogram was constructed based on risk scores and other clinical factors. HCC patients with a higher risk score showed a poor prognosis, and there were significant differences in immune cell infiltration between the high- and low-risk groups. In addition, three potential drugs associated with prognostic genes, streptozocin, norfloxacin, and hydrocotarnine, were identified. ***Conclusions:*** This study investigated the expression patterns of SGMGs and their relationship with tumor characteristics, resulting in the development of a novel model for predicting the prognosis of HCC patients. The study provides a reference for clinical prognosis prediction and treatment of HCC patients.

Citation of this article: Cai X, Xu F, Wang Z, Chen H, Lin S. Prognostic Biomarkers for Hepatocellular Carcinoma Based on Serine and Glycine Metabolism-related Genes. J Clin Transl Hepatol 2024;12(3):266–277. doi: 10.14218/JCTH.2023.00457.

Keywords: Serine and glycine metabolism-related genes; Hepatocellular carcinoma; Risk group; Prognosis; Tumor immune microenvironment.

Abbreviations: CTL, cytotoxic T lymphocyte; DCA, decision curve analysis; DEGs, differentially expressed genes; GEO, gene expression omnibus; HCC, hepatocellular carcinoma; IILs, immune infiltration levels; OS, overall survival; TCGA, The Cancer Genome Atlas; TIME, tumor immune microenvironment; SGMGs, serine and glycine metabolism-related genes.

[#]Contributed equally to this work.

[*]**Correspondence to:** Shengzhang Lin, Department of Clinical Medicine, School of Medicine, Hangzhou City University, 51 Huzhou Street, Gongshu District, Hangzhou, Zhejiang 310015, China. ORCID: https://orcid.org/0009-0000-9043-3712. Tel: +86-13605800785, E-mail: zdcsxylsz@163.com

Introduction

According to 2020 global cancer statistics Hepatocellular carcinoma (HCC) is one of the deadliest cancers, causing more than 830,000 deaths worldwide each year.[1] The occurrence of liver cancer is related to risk factors like hepatitis virus infection, excessive alcohol consumption, aflatoxin exposure, and genetic mutations.[2–5] HCC is currently treated with surgery, chemotherapy, radiotherapy, targeted therapy, immunotherapy, and liver transplantation.[6,7] Liver resection and transplantation are the primary treatment methods for patients with early-stage HCC.[8,9] However, surgery carries a high risk of recurrence, and matching suitable organ donors is difficult.[10] Therefore, exploring new treatment methods and screening for more effective prognostic markers are necessary and urgent to improve the prognosis of HCC patients.

Serine and glycine are essential amino acids in the human body that not only contribute to the formation of nucleic acids, lipids, amino acids, and coenzymes but also mediate numerous biological synthesis and signal transduction pathways.[11–13] Previous studies have shown that the synthesis and metabolism of serine and glycine are pivotal in tumor onset and development.[14,15] For example, they play a role in cancer cell nucleotide synthesis,[16] lipid metabolism,[17] and carbon source supplementing for cancer cell one-carbon metabolism.[12] Understanding the metabolic properties of serine and glycine may help to comprehend tumor progression. Recent studies on cancer metabolomics have found that increasing the synthesis of serine and glycine stimulates the synthesis of macromolecules in cancer cells, neutralizes high levels of oxidative stress, modulates methylation and tRNA acetylation, and promotes the occurrence and development of tumors.[16,18–21] Downregulation of the synthesis and metabolism of serine and glycine can effectively inhibit the proliferation of cancer cells.[22,23] In summary, the synthesis and metabolism of serine and glycine are closely linked to the occurrence and development of tumors. However, the corre-

lation between serine and glycine metabolism-related genes (SGMGs) and the prognosis of HCC, as well as their correlation with immune cells and the tumor immune microenvironment (TIME), is unclear and requires elucidation.

We developed and validated a novel prediction model based on SGMGs using bioinformatics. Additionally, we investigated the relationship between this model and tumor immune cell infiltration to effectively predict the overall survival (OS) rate of HCC patients and generate more accurate risk stratification management. Moreover, we used the TIDE score to analyze the immunotherapy effects of varying risk groups based on the risk model. Our findings can aid in the development of personalized immunotherapy for HCC patients.

Methods

Acquisition of SGMGs in HCC

We searched the LIHC dataset from The Cancer Genome Atlas (TCGA) database as the training cohort, which included clinical data and gene expression profiles from 424 (50 normal and 374 cancer) samples. In addition, we obtained the GSE76427 and GSE14520 datasets from the GEO database as external validation cohorts, which included mRNA expression profiles from 114 HCC samples. The clinical characteristics of patients in the data sets are detailed in Supplementary Table 1. The relevant literature yielded 24 SGMGs (Supplementary Table 2).[14] Using the R package edgeR,[24] we conducted differential analysis on cancer (n=375) and normal (n=32) samples in the training cohort (|logFC|>1.0, FDR<0.05) to identify differentially expressed genes (DEGs). Subsequently, we took the intersection of these DEGs and SGMGs to identify SGMGs that were differentially expressed in HCC.

Identification of different HCC subtypes based on SGMGs

Based on expression differences of SGMGs in HCC, we performed unsupervised clustering of the training cohort data using the ConsensusClusterPlus package, with 1000 repetitions. We conducted survival analysis on the clustered samples using the R package survival to compare the survival differences. Then, we performed differential analysis (|logFC|>1.0, FDR<0.05) on samples of different subtypes using the R package edgeR. Subsequently, we conducted KEGG and GO enrichment analyses on DEGs using the clusterProfiler R package and visualized the results with the enrichplot R package. Based on these DEGs, we constructed a PPI network in STRING, using interactions with a confidence score greater than 0.9 as the basis for constructing the PPI network.

Establishment of a prognostic model according to DEGs between different subtypes

Based on the DEGs in the PPI network, we established a prognostic model in the TCGA training cohort. Firstly, we identified genes that were significantly linked to survival through univariate Cox regression analysis. Then, we further narrowed the range of genes using the LASSO penalty method. During this process, we adjusted the optimal parameter λ through 10-fold cross-validation and calculated the optimal cutoff value using the "survminer" package. Subsequently, we performed multivariate Cox regression analysis on LASSO-selected genes to determine the final candidate genes and their risk coefficients. Finally, a risk score prognostic model was constructed.

We calculated the risk score value using the formula: Model=∑Coefficient(gene)*Expressionvalue(gene). Based on the median value, we assigned the training cohort into high- and low-risk groups. We then compared the survival rate differences using Kaplan-Meier survival curves, and assessed model sensitivity and specificity by ROC curves. These analyses were performed using the timeROC[25] and the survival packages. Additionally, we calculated and presented risk score values and corresponding survival outcomes for each patient in the training and validation cohorts. We plotted expression differences of feature genes between risk groups.

Independent testing of prognostic model for patients

Univariate and multivariate Cox regression analyses were applied to assess independent prognostic value. Based on the prognostic model and clinicopathological features of HCC, a nomogram was generated using the survival R package, regplot R package, and rms R package. To dissect the deviation between the nomogram and the ideal model, calibration curves for 1, 3, and 5-year survival rates were generated using the rms package. ROC curves and decision curve analysis (DCA) curves were plotted to assess the predictive ability of the model.

Analysis of TIME differences

CIBERSORT was used to calculate immune infiltration levels (IILs) of immune cells in high- and low-risk groups, and the Wilcox test was used to evaluate differences in TIME between different risk groups. TIDE is a computational method that predicts the immunotherapy response by simulating the main mechanisms of tumor immune escape: high cytotoxic T lymphocyte (CTL) levels induce T cell dysfunction, while low CTL levels inhibit T cell infiltration.[26] In this study, we predicted the response of tumor patients to immunotherapy by calculating the TIDE score. The Imvigor210 dataset included expression data, complete survival data, follow-up information, and immunotherapy response information for 348 cases of urothelial carcinoma downloaded from the Imvigor210CoreBiologies R package. GSE78220 was a melanoma dataset in which patients received anti-PD-1 checkpoint inhibition therapy. Patients in the Imvigor210 and GSE78220 cohorts were assigned into 2 groups according to their response to immunotherapy: response and non-response.

Tumor mutation analysis and heterogeneity analysis

The "maftools" package was used to assay and plot the mutation status of high- and low-risk groups for HCC SNV mutation data. The similarities and differences in mutation types, SNV class, and mutation rates between high- and low-risk groups were assessed, and the top ten genes with the highest mutation rates were selected. Additionally, waterfall plots were drawn to show the mutation status of feature genes.

Prediction of drugs related to feature genes in the model

To find new potential targets and more effective therapeutic drugs, we first downloaded drug/compound-related gene sets (http://dsigdb.tanlab.org/Downloads/DsigDB_All_detailed.txt) from the DsigDB database (http://tanlab.ucdenver.edu/DsigDB). Then, the clusterProfiler package and BiocFileCache package were used to perform drug prediction enrichment analysis on the feature genes in the model, and a network diagram and corresponding heat map of drugs and their enriched genes were plotted. Finally, potential small molecule compounds that were significantly associated with both sensitivity and prognostic genes were screened.

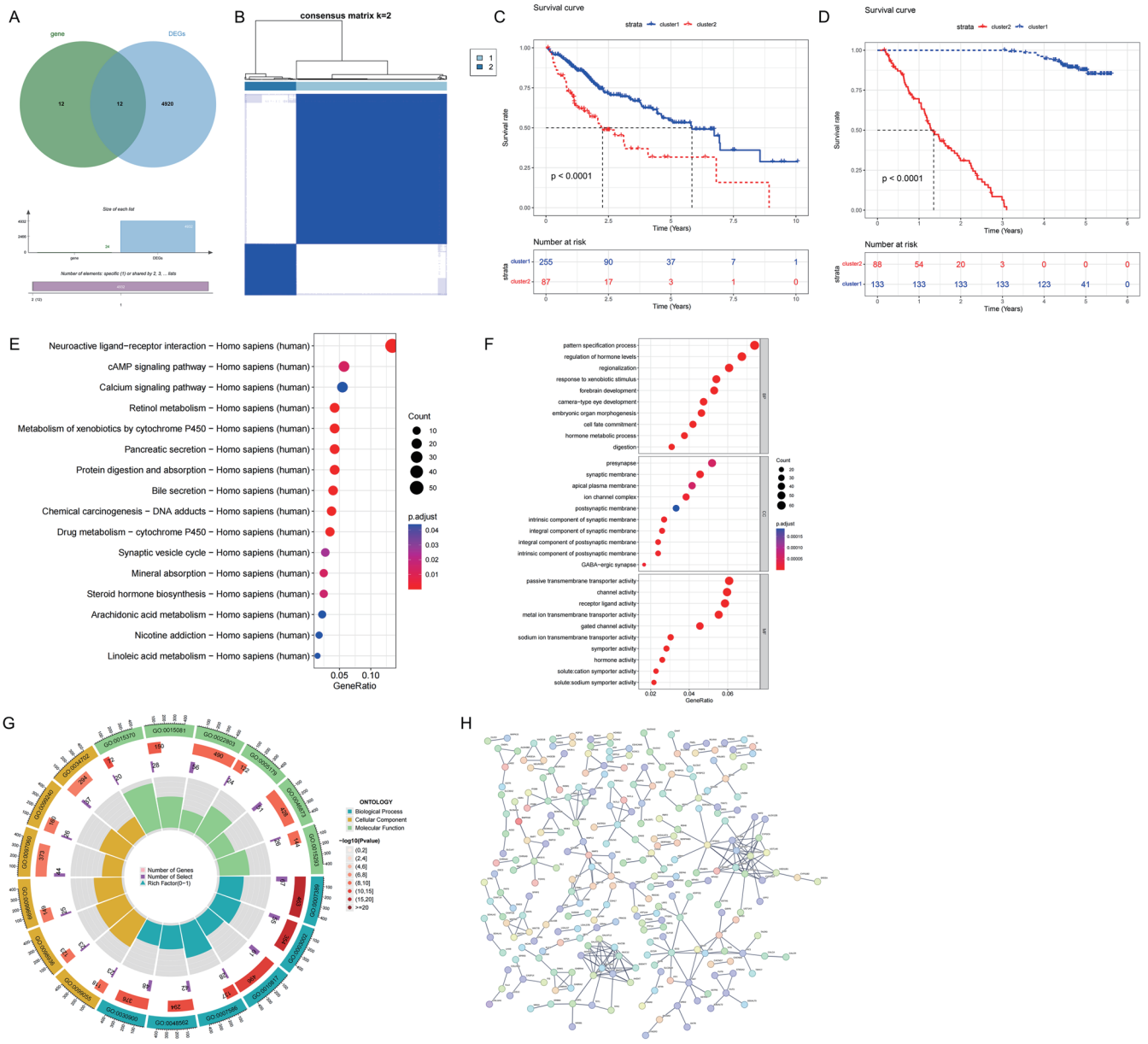

Fig. 1. Identification of differential SGMGs and different HCC subtypes. (A) Venn diagram showing the differentially expressed SGMGs in HCC. (B) Cluster analysis identifying different HCC subtypes. (C–D) Survival analysis exploring the survival differences between different subtypes in TCGA training set and GSE14520 validation set. (E) KEGG enrichment analysis exploring the enriched pathways of DEGs between subtypes. (F–G) Exploring the potential functions of DEGs between subtypes. (H) PPI network of DEGs between subtypes. DEGs, differentially expressed genes; HCC, hepatocellular carcinoma; KEGG, Kyoto Encyclopedia of Genes and Genomes; PPI, Protein-Protein Interaction; SGMGs, serine and glycine metabolism-related genes; TCGA, The Cancer Genome Atlas.

Results

Identification of differential SGMGs and different HCC subtypes

In the TCGA training cohort, we identified 4932 DEGs (3892 upregulated; 1040 downregulated) through differential analysis (Supplementary Table 3). Combining the results of literature retrieval, we identified 12 SGMGs that were differentially expressed in HCC by Venn analysis (Fig. 1A). Subsequently, we sorted the TCGA training cohort into 2 stable subtypes based on SGMG levels, with cluster1 containing 255 patients and cluster2 containing 87 patients (Fig. 1B). Survival analysis showed that cluster2 had a worse OS than cluster1 (Fig. 1C). After the patients in the GSE14520 validation cohort were divided into two clusters, the OS of cluster2 was also worse than that of cluster1 (Fig. 1D). In addition, compared to cluster2, we identified 1038 DEGs in cluster1. Enrichment analysis unraveled that DEGs were mainly enriched in KEGG pathways such as hormone metabolism, chemical carcinogenesis-DNA adducts by passive transmembrane transporters, and the cAMP signaling pathway (Fig. 1E). Besides, these DEGs were enriched in GO functions such as regulation of hormone levels, response to exogenous stimulus, passive transmembrane transporter

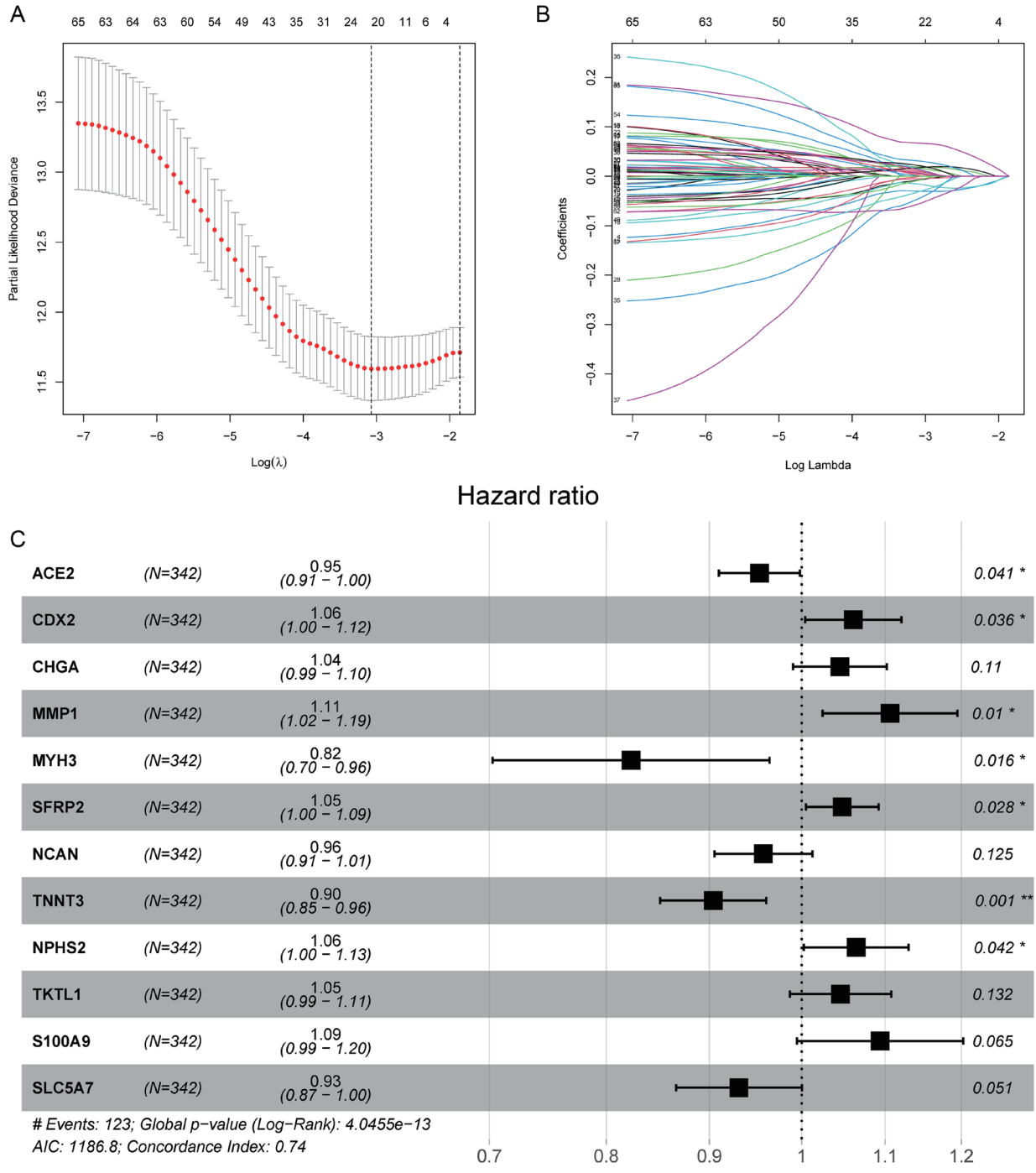

Fig. 2. Prognostic model construction and validation. (A) Coefficient distribution plot of LASSO Cox regression to eliminate multicollinearity. (B) Coefficient spectrum plot. (C) Forest plot of multivariate Cox regression analysis. *p<0.05; **p<0.01. LASSO, Least Absolute Shrinkage and Selection Operator.

activity, channel activity, and receptor ligand activity (Fig. 1F–G). According to the PPI network, these DEGs may have complex interactions (Fig. 1H). In conclusion, preliminary analysis revealed the role of SGMGs in HCC classification and the influence of SGMGs on HCC prognosis.

Prognostic model construction and validation

To further investigate the role of SGMGs in HCC prognosis, we developed an effective and clinically applicable robust risk feature model. Firstly, we used univariate Cox regression analysis in the TCGA training cohort to screen 109 survival-associated genes (threshold p<0.05). Then, we further screened out superior prognostic features from the 109 genes through LASSO Cox regression analysis (Fig. 2A–B). Finally, we used multivariate regression analysis to identify 12 crucial non-zero coefficient genes to construct a relevant risk

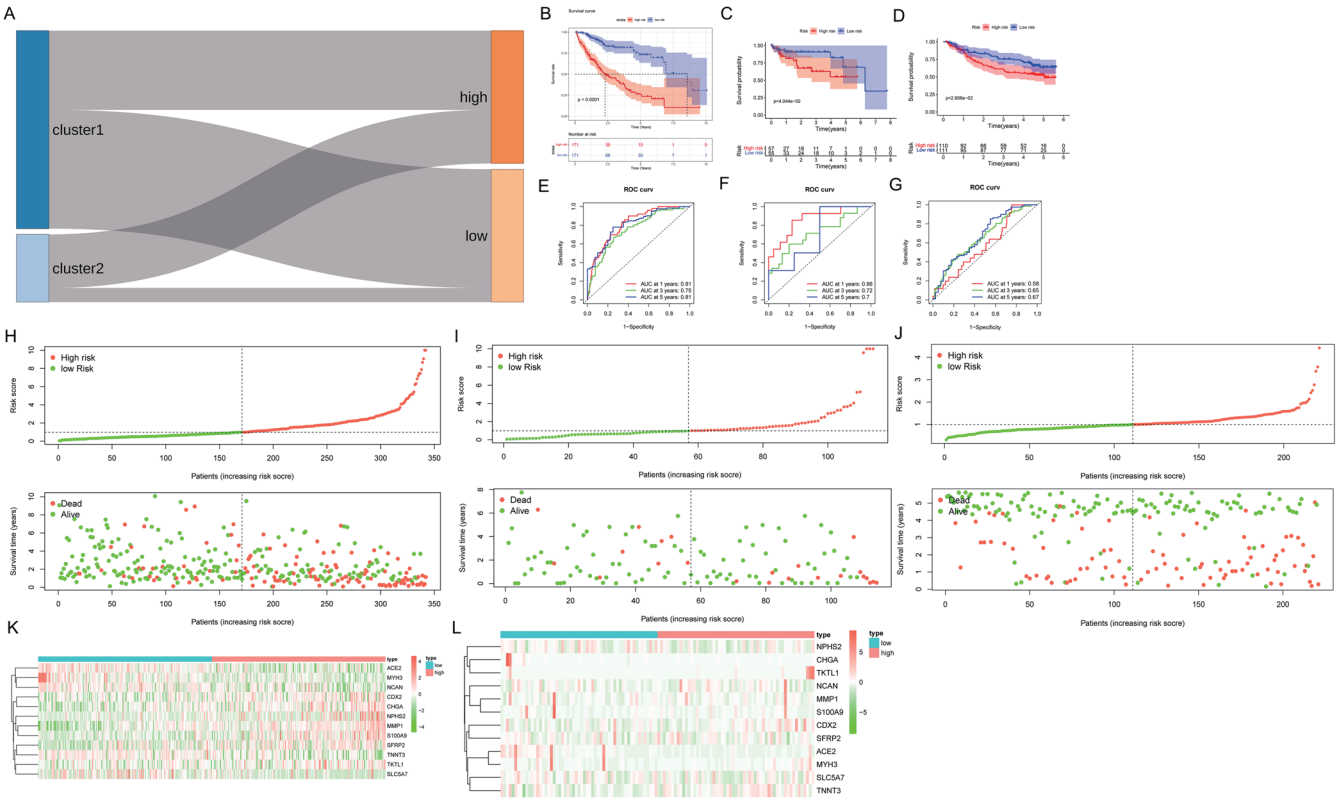

Fig. 3. Prognostic model construction and validation. (A) Sankey diagram summarizing the relationship between clustering and risk groups. (B) Survival curve of training set cohort. (C–D) Survival curve of the validation set cohort. (E) ROC curve of training set cohort. (F–G) ROC curve of the validation set cohort. (H) Scatter plot of risk score and survival status distribution in the training set cohort. (I–J) Scatter plot of risk score and survival status distribution in the validation set cohort. (K) Heat map of expression levels of feature genes in the training set cohort of the risk score prognosis model. (L) Heat map of expression levels of feature genes in the validation set cohort of the risk score prognosis model. ROC, Receiver Operating Characteristic; AUC, Area Under Curve.

score model. Among them, MMP1, S100A9, NPHS2, CDX2, SFRP2, TKTL1, and CHGA were identified as prognostic risk factors, while NCAN, ACE2, SLC5A7, TNNT3, and MYH3 were identified as prognostic protective factors (Fig. 2C). Notably, CHGA, NCAN, TKTL1, S100A9, and SLC5A7 seemed to be statistically insignificant ($p>0.05$), and we speculated that they may be associated with other markers and outcomes.

Regarding survival analysis, we computed risk scores using a formula. The cases in the TCGA training cohort as well as external validation cohorts GSE76427 and GSE14520 were assigned into high- and low-risk groups according to the median risk score. The Sankey diagram summarized the relationship between clustering and risk grouping (Fig. 3A). In different datasets, Kaplan-Meier survival curves demonstrated differences in survival rates between different risk groups. It could be observed that the low-risk group exhibited better OS in the training and validation cohorts (Fig. 3B–D). The average AUC values for 1, 3, and 5-year prognosis prediction in the TCGA cohort were 0.81, 0.75, and 0.81, respectively (Fig. 3E). The average AUC values for 1, 3, and 5-year survival rates were 0.88, 0.72, and 0.70 in GSE76427, and 0.58, 0.65, 0.67 in GSE14520 (Fig. 3F–G). Additionally, scatter plots showed the distribution of risk scores and survival status in both datasets (Fig. 3H–J). The expression of feature genes in the prognostic model was presented in the form of a heatmap. Among them, ACE2, MYH3, NCAN, TNNT3, and SLC5A7 were highly expressed in the low-risk group, while MMP1, S100A9, NPHS2, CDX2, SFRP2, TKTL1, and CHGA were highly expressed in the high-risk group (Fig. 3K–L). These analyses demonstrated that the risk score was capable of reliably differentiating tumor prognosis differences.

Independent evaluation of prognostic model

To investigate the independence of the prognostic model in predicting patient survival, we performed univariate and multivariate Cox regression analyses based on the clinical factors and risk scores of each sample in the TCGA dataset (Fig. 4A–B). We found that the risk score could serve as an independent prognostic factor in predicting the prognosis of HCC patients. We combined the risk score with prognostic clinical features and constructed a nomogram to predict 1, 3, and 5-year OS (Fig. 4C). The calibration chart indicated that our constructed nomogram performed well compared to the ideal model (Fig. 4D). DCA revealed that the nomogram had favorable predictive performance (Fig. 4E). ROC curves presented that the joint model had the highest AUC value (Fig. 4F), demonstrating that the joint model had the best prognostic predictive efficiency. Taken together, these results indicated that we had successfully constructed a reliable prognostic model with high predictive value.

Evaluation of TIME and immune cell IILs in different risk groups

Considering the relationship between SGMGs and immunity, we applied CIBERSORT to assess the IILs of each HCC patient. The results showed that B cell naïve, B cell memo-

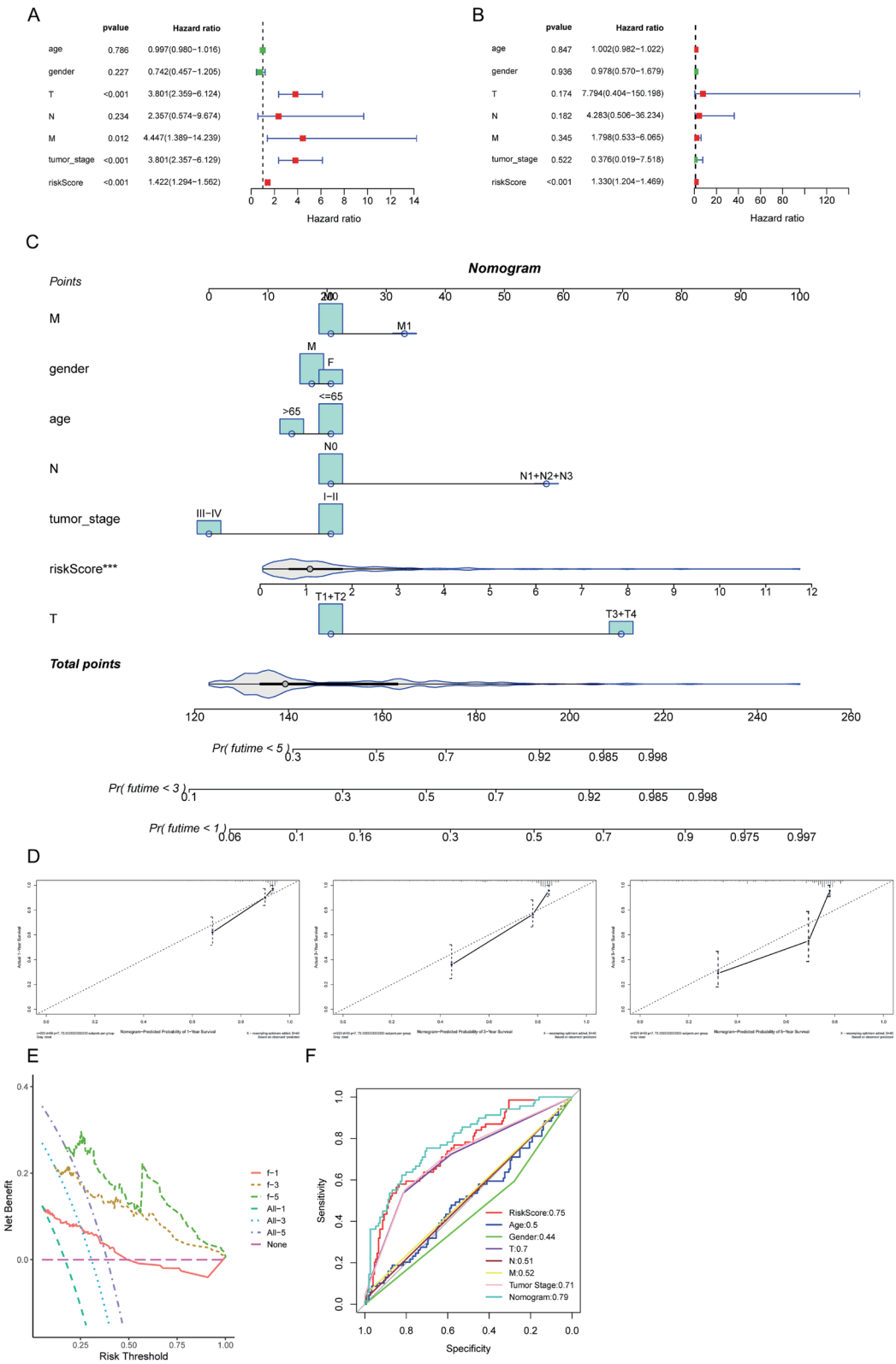

Fig. 4. Independent evaluation of prognostic model. (A) Univariate analysis of clinical features and risk score. (B) Multivariate analysis of clinical features and risk score. (C) Nomogram constructed based on clinical features and risk score. (D) Calibration curve of 1/3/5-year correction curve of the nomogram. (E) DCA curve of the nomogram. (F) ROC curve of clinical features, RiskScore, and nomogram. ***p<0.001. DCA, decision curve analysis; ROC, Receiver Operating Characteristic; T, Tumor; N, Node; M, metastasis.

ry, Monocytes, Dendritic cells activated, Mast cells resting, and Neutrophils all had higher IILs in the low-risk group, whereas T cell regulatory, Macrophages M0, and Dendritic cells resting had higher IILs in the high-risk group (Fig. 5A). Furthermore, to predict the likelihood of HCC patients responding to immunotherapy, we found that the low-risk group had more immune-responsive patients and a higher Response Percentage (Fig. 5B). In addition, given that patients in varying groups showed varying IILs, we speculated that there may be some differences in the efficacy of ICI treatment between the two groups. Since transcriptome data for HCC patients receiving ICIs were unavailable, data from other malignancies were utilized to support this theory. In order to confirm the link between risk score and immunological response, we examined the IMvigor210 dataset. We discovered that the Response Percentage was greater in the low-risk group (Fig. 5C). At the same time, OS was noticeably better in the low-risk group (Fig. 5D). Additionally, we confirmed the association between risk score and immune response using the GSE78220 dataset, and discovered that Response Percentage was greater in the low-risk group (Fig. 5E). At the same time, OS was notably better in the low-risk group (Fig. 5F). These results suggested that immunotherapy was more effective in low-risk patients. This could explain why HCC patients with low risk had better survival than those with high risk.

Comparing mutation status between high-risk and low-risk groups

Cancer is thought to occur most frequently as a result of gene alterations. To dissect the potential mechanism of using risk score features to evaluate patient prognosis, we studied the gene mutations of patients in different risk groups. Missense mutations had the highest proportion in both groups, followed by frameshift mutations. SNP mutation was the most prevalent type, with C>T and C>A accounting for 50.4% and 46.2% of SNP mutations in high- and low-risk groups, respectively (Fig. 6A–B). The stacked bar chart showed the top 10 mutated genes, with TP53, TTN, MUC16, CTNNB1, RYR2, ALB, PCLO, ARID1A, APOB, and OBSCN in the high-risk group, and TTN, CTNNB1, TP53, MUC16, ALB, PCLO, ABCA13, XIRP2, AXIN1, and APOB in the low-risk group (Fig. 6A–B). For model genes, CHGA, NCAN, and MYH3 were the top 3 mutated genes in the high-risk group, while SLC5A7, ACE2, and CDX2 were the top 3 mutated genes in the low-risk group (Fig. 6C–D).

Drug sensitivity prediction

Furthermore, using the DSigDB database, we selected some potential drugs that were significantly related to both sensitivity and prognosis genes. Streptozocin, norfloxacin, and hydrocotarnine were the top three significantly enriched small molecule compounds (Fig. 7A). In addition, enrichment analysis of target genes showed that streptozocin was enriched in three feature genes (ACE2, NPHS2, and MMP1), while norfloxacin and hydrocotarnine were enriched in two feature genes (Fig. 7B–C). MMP1 was the feature gene that was commonly enriched by all three small molecule compounds, indicating that MMP1 may be a key gene for the action of these potential small molecule compounds.

Discussion

Serine and glycine metabolism is a non-glutamine amino acid metabolism component that is important in energy metabolism. This metabolic pathway has been shown to be critical for cancer cell survival by participating in energy production, biosynthesis, and signal transduction during tumor occurrence and development.[13,15,27] As HCC is a molecularly heterogeneous malignancy, its molecular characteristics are linked to tumor biology behavior.[28] Therefore, identifying molecular biomarkers linked to serine and glycine metabolism in HCC is of great significance.

Currently, although many researchers have established numerous prognostic models for HCC,[29–31] there is no prognostic model related to SGMGs to systematically predict and evaluate the overall prognosis of HCC patients. Therefore, to reveal the potential function and unique prognostic value of SGMGs in HCC, we examined public datasets through bioinformatics analysis to provide new insights for future research.

We successfully built a 12-gene risk score model, in which ACE2, MYH3, NCAN, TNNT3, and SLC5A7 were upregulated in the low-risk group, while MMP1, S100A9, NPHS2, CDX2, SFRP2, TKTL1, and CHGA were upregulated in the high-risk group of HCC. MMP1 is a downstream member of TCONS_00012883 and has been shown to mediate the progression of colorectal cancer with DDX3, YY1, and TCONS_00012883.[32] In addition, MMP1 promotes metastasis of ovarian cancer and esophageal squamous cell carcinoma.[33,34] S100A9 is a widely studied biomarker in recent years and a pivotal mediator of cancer progression.[35–37] It is worth noting that clinical studies have found that downregulation of S100A9 expression in HCC can significantly inhibit cancer cell growth.[38] Therefore, targeting S100A9 is a possible strategy for HCC patients. CDX2 is a critical prognostic biomarker for patients with stage I and II colon cancer.[39] CDX2 is expressed positively in HCC,[40–42] congruous with the findings of this study. Overexpression of SFRP2, TKTL1, and CHGA is implicated in a dismal prognosis in various cancers. For example, Lai *et al*.'s study[43] found that increased SFRP2 was substantially linked to invasive features of urothelial carcinoma and could lead to a poor prognosis. TKTL1 has been shown to be overexpressed in malignant conjunctival tumors and is associated with clinical outcomes (e.g., tumor relapse rate).[44] In addition, Allison B Weisbrod *et al*. 's study[45] found that changes in CHGA expression are related to the phenotype of invasive pancreatic neuroendocrine tumors and can be a useful prognostic marker. Combining the findings of this study, the risk score composed of these genes can be an independent prognosticator for predicting the prognosis of HCC patients.

Furthermore, this study found marked differences in the immune landscape of patients in different risk groups. This may suggest that different expression levels of SGMGs could impact the TIME of HCC patients. He *et al.*[46] found that enhancing extracellular levels of serine significantly inhibited macrophage and neutrophil function. Su *et al.*[47] found that serine and glycine synthesis switch macrophage phenotypes to express immune checkpoint molecule PD-L1 by inducing IL-1β production. Similarly, in this study, we found that macrophages M0 had higher IILs in the high-risk group, while neutrophils had higher IILs in the low-risk group. This finding is not entirely consistent with previous studies. Combining the other results of this study, we found that in addition to macrophages and neutrophils, regulatory T cells, resting dendritic cells, naive B cells, memory B cells, monocytes, and resting mast cells also had different IILs between the two groups. Therefore, we speculated that differences in TIME between varying risk groups may be the result of the combined action of multiple immune cells. However, the molecular mechanisms underlying this require further investigation. Thus, serine and glycine metabolism not only participate in tumorigenesis but also have a connection to TIME.

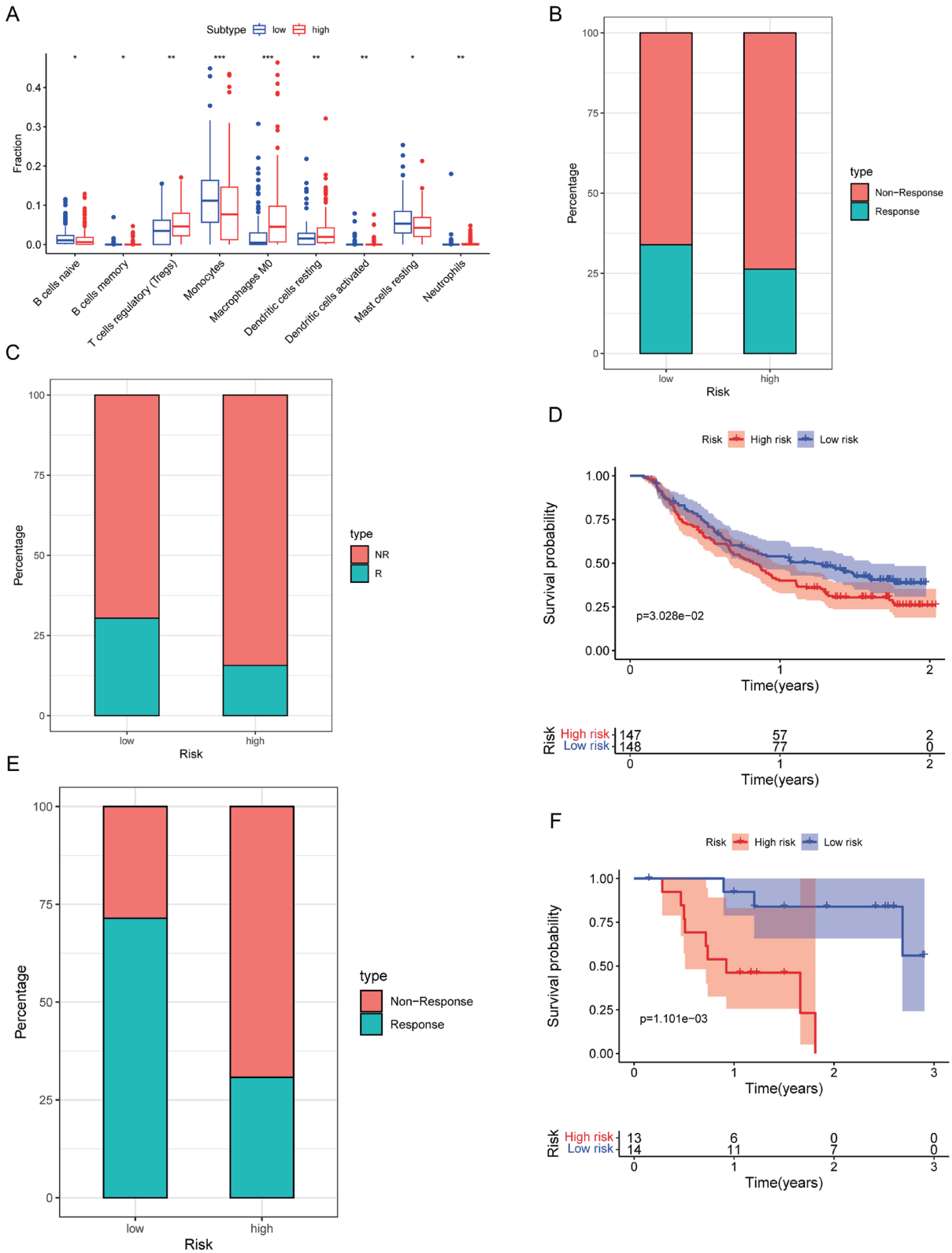

Fig. 5. Evaluation of TIME and immune cell IILs in different risk groups. (A) CIBERSORT exploring the differences in IILs between different risk groups. (B) TIDE predicting the immunotherapy response of HCC patients. (C) ICI treatment response in IMvigor210. (D) Survival curve of high- and low-risk groups in IMvigor210. (E) ICI treatment response in GSE78220. (F) Survival curve of high- and low-risk groups in GSE78220. *p<0.05, **p<0.01, ***p<0.001. ICI, Immune checkpoint inhibitors; IILs, immune infiltration levels; TIDE, Tumor Immune Dysfunction and Exclusion; TIME, tumor immune microenvironment.

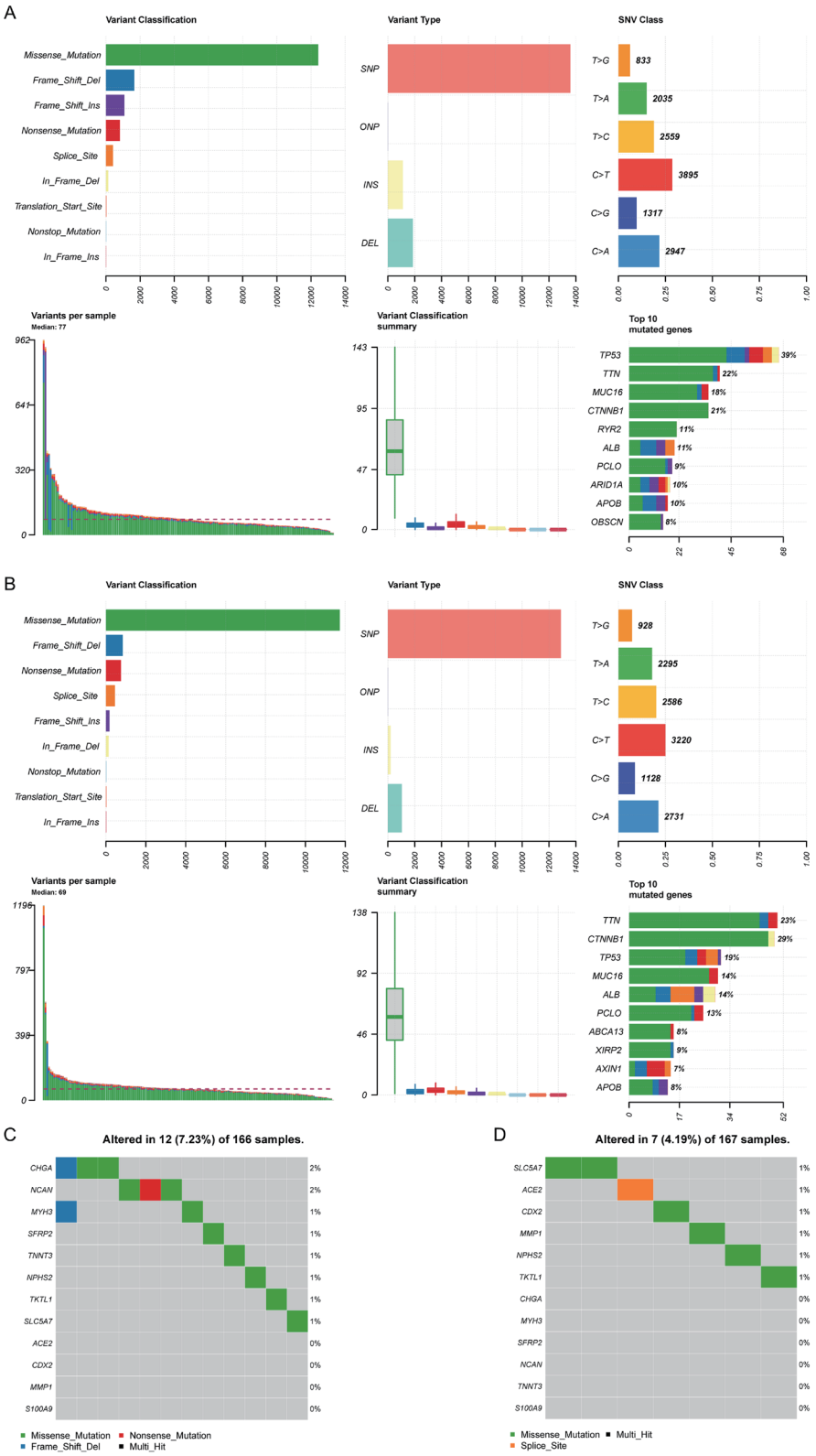

Fig. 6. Comparing mutation status between high-risk and low-risk groups. (A–B) Mutation landscape of high-risk (A) and low-risk (B) groups, including mutation classification statistics, mutation type statistics, subdivided statistics of single nucleotide polymorphism mutations, mutation number and proportion statistics of all samples, number statistics of different mutation types, and statistics of genes with the most mutations. (C–D) Model gene statistics of the top ten mutation rates in high-risk (C) and low-risk (D) groups. A, ADENINE; C, CYTOSINE; G, GUANINE; T, THYMINE.

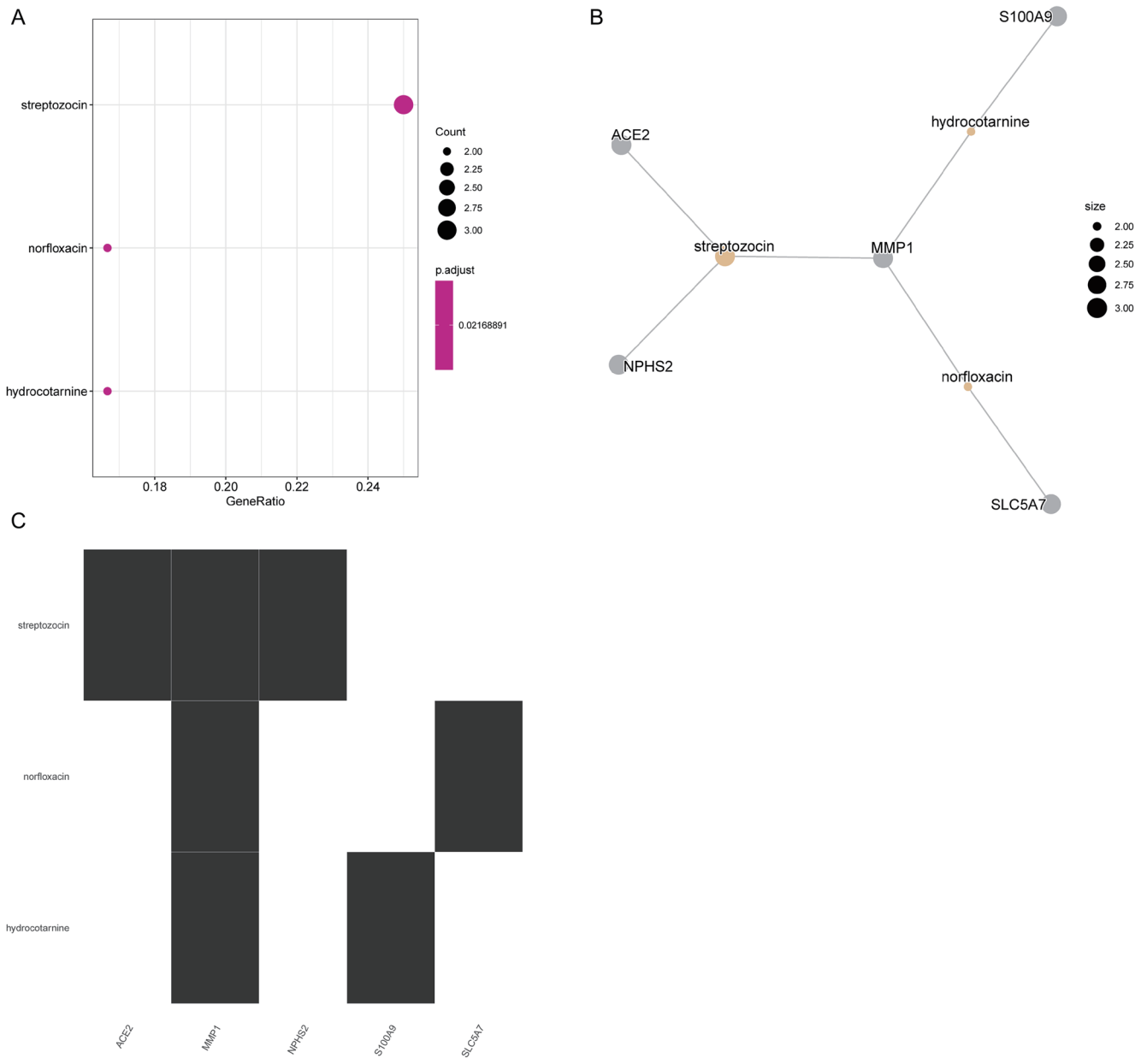

Fig. 7. Drug sensitivity prediction. (A) Enrichment analysis showing potential small molecule drugs screened based on the DSigDB database. (B) Network diagram showing the potential interaction between prognosis feature genes and screened small molecule drugs. (C) Enrichment drug results correspond to the prognosis genes.

We screened for potential drugs significantly associated with sensitivity and prognostic genes based on the DSigDB database. Although *in vitro* and *in vivo* experiments were not performed to validate results, enrichment analysis revealed that streptozocin was significantly enriched with three feature genes (ACE2, NPHS2, and MMP1) in the model, while norfloxacin and hydrocotarnine were significantly enriched with two feature genes each. Therefore, we speculated that these small molecule drugs may have possible therapeutic effects on HCC. Streptozocin can effectively inhibit tumor cell proliferation.[48] Current studies have shown that streptozocin can be widely used in the clinical treatment of pancreatic cancer,[49,50] but its role in HCC is still unknown.

We comprehensively assessed RNA sequencing data from HCC patient cohorts from TCGA and GEO, established a risk prognostic model for HCC based on SGMGs, investigated the correlation between SGMGs and the immune landscape of HCC, and predicted the possibility of immunotherapy for patients in different risk groups. Univariate and multivariate Cox regression analyses reported that the risk score model was an independent prognostic factor for HCC. In conclusion, this model had favorable predictive ability for the prognosis of HCC patients, providing the possibility for further improving personalized treatment for HCC patients. However, there are still limitations. Firstly, clinical data collected by bioinformatics analysis are limited and retrospective, and the specificity expression and function of SGMGs in HCC should be validated in prospective designs. Furthermore, the rela-

tionship between prognostic model-related genes and HCC is predicted in the bioinformatics analysis stage, which requires further molecular biology experiments for validation. Further research is needed to confirm the efficacy of the potential drugs screened in this study.

Funding

This work was supported by Postgraduate Science Foundation Project of Zhejiang Chinese Medical University (No. Y202248712).

Conflict of interest

The authors have no conflict of interests related to this publication.

Author contributions

Conception and design: XC. Administrative support: SL. Collection and assembly of data: FX, ZW. Data analysis and interpretation: HC, SL. Manuscript writing: XC, FX. Final approval of manuscript: All authors

Ethical statement

This study was carried out in accordance with the Declaration of Helsinki. The individual consent for this retrospective analysis was waived.

Data sharing statement

The data and materials in the current study are available from the corresponding author on reasonable request.

References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, *et al*. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71(3):209–249. doi:10.3322/caac.21660, PMID:33538338.
[2] Gao Q, Zhu H, Dong L, Shi W, Chen R, Song Z, *et al*. Integrated Proteogenomic Characterization of HBV-Related Hepatocellular Carcinoma. Cell 2019;179(5):1240. doi:10.1016/j.cell.2019.10.038, PMID:31730861.
[3] Rodriguez M, Gonzalez-Dieguez ML, Varela M, Cadahia V, Andres-Vizan SM, Mesa A, *et al*. Impact of Alcohol Abstinence on the Risk of Hepatocellular Carcinoma in Patients With Alcohol-Related Liver Cirrhosis. Am J Gastroenterol 2021;116(12):2390–2398. doi:10.14309/ajg.0000000000001399, PMID:34569986.
[4] Zhu Q, Ma Y, Liang J, Wei Z, Li M, Zhang Y, *et al*. AHR mediates the aflatoxin B1 toxicity associated with hepatocellular carcinoma. Signal Transduct Target Ther 2021;6(1):299. doi:10.1038/s41392-021-00713-1, PMID:34373448.
[5] Nault JC, Calderaro J, Di Tommaso L, Balabaud C, Zafrani ES, Bioulac-Sage P, *et al*. Telomerase reverse transcriptase promoter mutation is an early somatic genetic alteration in the transformation of premalignant nodules in hepatocellular carcinoma on cirrhosis. Hepatology 2014;60(6):1983–1992. doi:10.1002/hep.27372, PMID:25123086.
[6] Su GL, Altayar O, O'Shea R, Shah R, Estfan B, Wenzell C, *et al*. AGA Clinical Practice Guideline on Systemic Therapy for Hepatocellular Carcinoma. Gastroenterology 2022;162(3):920–934. doi:10.1053/j.gastro.2021.12.276, PMID:35210014.
[7] Ganesan P, Kulik LM. Hepatocellular Carcinoma: New Developments. Clin Liver Dis 2023;27(1):85–102. doi:10.1016/j.cld.2022.08.004, PMID:36400469.
[8] Chu KJ, Kawaguchi Y, Wang H, Jiang XQ, Hasegawa K. Update on the Diagnosis and Treatment of Combined Hepatocellular Cholangiocarcinoma. J Clin Transl Hepatol 2024;12(2):210–217. doi:10.14218/JCTH.2023.00189.
[9] Cescon M, Cucchetti A, Ravaioli M, Pinna AD. Hepatocellular carcinoma locoregional therapies for patients in the waiting list. Impact on transplantability and recurrence rate. J Hepatol 2013;58(3):609–618. doi:10.1016/j.jhep.2012.09.021, PMID:23041304.
[10] Merchant N, David CS, Cunningham SC. Early Hepatocellular Carcinoma: Transplantation versus Resection: The Case for Liver Resection. Int J Hepatol 2011;2011:142085. doi:10.4061/2011/142085, PMID:21994848.
[11] Sullivan MR, Vander Heiden MG. When cancer needs what's non-essential. Nat Cell Biol 2017;19(5):418–420. doi:10.1038/ncb3523, PMID:28422952.
[12] Locasale JW. Serine, glycine and one-carbon units: cancer metabolism in full circle. Nat Rev Cancer 2013;13(8):572–583. doi:10.1038/nrc3557, PMID:23822983.
[13] Pan S, Fan M, Liu Z, Li X, Wang H. Serine, glycine and one-carbon metabolism in cancer (Review). Int J Oncol 2021;58(2):158–170. doi:10.3892/ijo.2020.5158, PMID:33491748.
[14] Chen S, Zhang S, Feng W, Li J, Yuan Y, Li W, *et al*. Serine and glycine metabolism-related gene expression signature stratifies immune profiles of brain gliomas, and predicts prognosis and responses to immunotherapy. Front Pharmacol 2022;13:1072253. doi:10.3389/fphar.2022.1072253, PMID:36467068.
[15] Jain M, Nilsson R, Sharma S, Madhusudhan N, Kitami T, Souza AL, *et al*. Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. Science 2012;336(6084):1040–1044. doi:10.1126/science.1218595, PMID:22628656.
[16] Fan TWM, Bruntz RC, Yang Y, Song H, Chernyavskaya Y, Deng P, *et al*. De novo synthesis of serine and glycine fuels purine nucleotide biosynthesis in human lung cancer tissues. J Biol Chem 2019;294(36):13464–13477. doi:10.1074/jbc.RA119.008743, PMID:31337706.
[17] Gao X, Lee K, Reid MA, Sanderson SM, Qiu C, Li S, *et al*. Serine Availability Influences Mitochondrial Dynamics and Function through Lipid Metabolism. Cell Rep 2018;22(13):3507–3520. doi:10.1016/j.celrep.2018.03.017, PMID:29590619.
[18] Geeraerts SL, Heylen E, De Keersmaecker K, Kampen KR. The ins and outs of serine and glycine metabolism in cancer. Nat Metab 2021;3(2):131–141. doi:10.1038/s42255-020-00329-9, PMID:33510397.
[19] Newman AC, Maddocks ODK. One-carbon metabolism in cancer. Br J Cancer 2017;116(12):1499–1504. doi:10.1038/bjc.2017.118, PMID:28472819.
[20] DeBerardinis RJ. Serine metabolism: some tumors take the road less traveled. Cell Metab 2011;14(3):285–386. doi:10.1016/j.cmet.2011.08.004, PMID:21907134.
[21] Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, *et al*. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. Nature 2009;457(7231):910–914. doi:10.1038/nature07762, PMID:19212411.
[22] Mullarky E, Lucki NC, Beheshti Zavareh R, Anglin JL, Gomes AP, Nicolay BN, *et al*. Identification of a small molecule inhibitor of 3-phosphoglycerate dehydrogenase to target serine biosynthesis in cancers. Proc Natl Acad Sci U S A 2016;113(7):1778–1783. doi:10.1073/pnas.1521548113, PMID:26831078.
[23] Pacold ME, Brimacombe KR, Chan SH, Rohde JM, Lewis CA, Swier LJ, *et al*. A PHGDH inhibitor reveals coordination of serine synthesis and one-carbon unit fate. Nat Chem Biol 2016;12(6):452–458. doi:10.1038/nchembio.2070, PMID:27110680.
[24] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26(1):139–140. doi:10.1093/bioinformatics/btp616, PMID:19910308.
[25] Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. Stat Med 2013;32(30):5381–5897. doi:10.1002/sim.5958, PMID:24027076.
[26] Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, *et al*. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med 2018;24(10):1550–1558. doi:10.1038/s41591-018-0136-1, PMID:30127393.
[27] Sowers ML, Herring J, Zhang W, Tang H, Ou Y, Gu W, *et al*. Analysis of glucose-derived amino acids involved in one-carbon and cancer metabolism by stable-isotope tracing gas chromatography mass spectrometry. Anal Biochem 2019;566:1–9. doi:10.1016/j.ab.2018.10.026, PMID:30409761.
[28] Erstad DJ, Fuchs BC, Tanabe KK. Molecular signatures in hepatocellular carcinoma: A step toward rationally designed cancer therapy. Cancer 2018;124(15):3084–3104. doi:10.1002/cncr.31257, PMID:29663340.
[29] He D, Cai L, Huang W, Weng Q, Lin X, You M, *et al*. Prognostic value of fatty acid metabolism-related genes in patients with hepatocellular carcinoma. Aging (Albany NY) 2021;13(13):17847–17863. doi:10.18632/aging.203288, PMID:34257161.
[30] Xu Y, Zhang Z, Xu D, Yang X, Zhou L, Zhu Y. Identification and integrative analysis of ACLY and related gene panels associated with immune microenvironment reveal prognostic significance in hepatocellular carcinoma. Cancer Cell Int 2021;21(1):409. doi:10.1186/s12935-021-02108-2, PMID:34344378.
[31] Zhang H, Xia P, Liu J, Chen Z, Ma W, Yuan Y. ATIC inhibits autophagy in hepatocellular cancer through the AKT/FOXO3 pathway and serves as a prognostic signature for modeling patient survival. Int J Biol Sci 2021;17(15):4442–4458. doi:10.7150/ijbs.65669, PMID:34803509.
[32] Xie J, Zhong Y, Chen R, Li G, Luo Y, Yang J, *et al*. Serum long non-coding RNA LINC00887 as a potential biomarker for diagnosis of renal cell carcinoma. FEBS Open Bio 2020;10(9):1802–1809. doi:10.1002/2211-5463.12930, PMID:32654370.
[33] Yokoi A, Yoshioka Y, Yamamoto Y, Ishikawa M, Ikeda SI, Kato T, *et al*. Malignant extracellular vesicles carrying MMP1 mRNA facilitate peritoneal dissemination in ovarian cancer. Nat Commun 2017;8:14470. doi:10.1038/ncomms14470, PMID:28262727.
[34] Liu M, Hu Y, Zhang MF, Luo KJ, Xie XY, Wen J, *et al*. MMP1 promotes tumor growth and metastasis in esophageal squamous cell carcinoma. Cancer Lett 2016;377(1):97–104. doi:10.1016/j.canlet.2016.04.034, PMID:27130665.
[35] Chung YH, Volckaert BA, Steinmetz NF. Metastatic Colon Cancer Treatment Using S100A9-Targeted Cowpea Mosaic Virus Nanoparticles. Biomacromolecules 2022;23(12):5127–5136. doi:10.1021/acs.biomac.2c00873,

PMID:36375170.
[36] Biswas AK, Han S, Tai Y, Ma W, Coker C, Quinn SA, *et al*. Targeting S100A9-ALDH1A1-Retinoic Acid Signaling to Suppress Brain Relapse in EGFR-Mutant Lung Cancer. Cancer Discov 2022;12(4):1002–1021. doi:10.1158/2159-8290.CD-21-0910, PMID:35078784.
[37] Symanowski JT, Kim ES. Gene expression and prognosis in bladder cancer—real progress? Editorial on 'S100A9 and EGFR gene signatures predict disease progression in muscle invasive bladder cancer patients after chemotherapy'. Ann Oncol 2014;25(5):919–920. doi:10.1093/annonc/mdu113, PMID:24608197.
[38] Zhong C, Niu Y, Liu W, Yuan Y, Li K, Shi Y, *et al*. S100A9 Derived from Chemoembolization-Induced Hypoxia Governs Mitochondrial Function in Hepatocellular Carcinoma Progression. Adv Sci (Weinh) 2022;9(30):e2202206. doi:10.1002/advs.202202206, PMID:36041055.
[39] Dalerba P, Sahoo D, Paik S, Guo X, Yothers G, Song N, *et al*. CDX2 as a Prognostic Biomarker in Stage II and Stage III Colon Cancer. N Engl J Med 2016;374(3):211–222. doi:10.1056/NEJMoa1506597, PMID:26789870.
[40] Kmeid M, Lukose G, Hodge K, Cho D, Kim KA, Lee H. Aberrant expression of SATB2, CDX2, CDH17 and CK20 in hepatocellular carcinoma: a pathological, clinical and outcome study. Histopathology 2021;79(5):768–778. doi:10.1111/his.14420, PMID:34036629.
[41] El Jabbour T, Durie N, Lee H. Coexpression of CDX2 and CK20 in hepatocellular carcinoma, an exceedingly rare co-incidence with potential diagnostic pitfall. Hum Pathol 2018;81:298. doi:10.1016/j.humpath.2018.04.031, PMID:29953895.
[42] Chandan VS, Shah SS, Torbenson MS, Wu TT. Co-expression of CDX2 and CK20 in hepatocellular carcinoma, an exceedingly rare co-incidence with potential diagnostic pitfall-reply. Hum Pathol 2018;81:299. doi:10.1016/j.humpath.2018.04.030, PMID:29981308.
[43] Lai HY, Chiu CC, Kuo YH, Tsai HH, Wu LC, Tseng WH, *et al*. High Stromal SFRP2 Expression in Urothelial Carcinoma Confers an Unfavorable Prognosis. Front Oncol 2022;12:834249. doi:10.3389/fonc.2022.834249, PMID:35372028.
[44] Lange CA, Tisch-Rottensteiner J, Bohringer D, Martin G, Schwartzkopff J, Auw-Haedrich C. Enhanced TKTL1 expression in malignant tumors of the ocular adnexa predicts clinical outcome. Ophthalmology 2012;119(9):1924–1929. doi:10.1016/j.ophtha.2012.03.037, PMID:22658715.
[45] Weisbrod AB, Zhang L, Jain M, Barak S, Quezado MM, Kebebew E. Altered PTEN, ATRX, CHGA, CHGB, and TP53 expression are associated with aggressive VHL-associated pancreatic neuroendocrine tumors. Horm Cancer 2013;4(3):165–175. doi:10.1007/s12672-013-0134-1, PMID:23361940.
[46] He F, Yin Z, Wu C, Xia Y, Wu M, Li P, *et al*. l-Serine Lowers the Inflammatory Responses during Pasteurella multocida Infection. Infect Immun 2019;87(12):e00677–19. doi:10.1128/IAI.00677-19, PMID:31570555.
[47] Su S, Zhao J, Xing Y, Zhang X, Liu J, Ouyang Q, *et al*. Immune Checkpoint Inhibition Overcomes ADCP-Induced Immunosuppression by Macrophages. Cell 2018;175(2):442–457.e23. doi:10.1016/j.cell.2018.09.007, PMID:30290143.
[48] Ducreux M, Dahan L, Smith D, O'Toole D, Lepere C, Dromain C, *et al*. Bevacizumab combined with 5-FU/streptozocin in patients with progressive metastatic well-differentiated pancreatic endocrine tumours (BETTER trial)—a phase II non-randomised trial. Eur J Cancer 2014;50(18):3098–3106. doi:10.1016/j.ejca.2014.10.002, PMID:25454412.
[49] Kouvaraki MA, Ajani JA, Hoff P, Wolff R, Evans DB, Lozano R, *et al*. Fluorouracil, doxorubicin, and streptozocin in the treatment of patients with locally advanced and metastatic pancreatic endocrine carcinomas. J Clin Oncol 2004;22(23):4762–4771. doi:10.1200/JCO.2004.04.024, PMID:15570077.
[50] Rogers JE, Lam M, Halperin DM, Dagohoy CG, Yao JC, Dasari A. Fluorouracil, Doxorubicin with Streptozocin and Subsequent Therapies in Pancreatic Neuroendocrine Tumors. Neuroendocrinology 2022;112(1):34–42. doi:10.1159/000514339, PMID:33434908.